Validation Studies using an Alloyed Gold Standard

Sholom Wacholder,[1] Ben Armstrong,[2] and Patricia Hartge[1]

A key assumption made when using a validation study to correct an estimate of relative risk for bias due to misclassification or measurement error is that the available measure, known to have error but nonetheless used routinely in the main study, is compared to a gold standard measured without error. In most epidemiologic applications, the putative gold standard is in fact measured with error. The effect of the violation of the assumption on the corrected estimate depends on the magnitudes of the errors in the two measures and on their correlation. In particular, when the errors are negatively correlated, independent, or weakly positively correlated, the corrected estimate will tend to overcorrect beyond the true value. *Am J Epidemiol* 1993;137:1251–8.

bias (epidemiology); biometry; misclassification; prospective study; retrospective study

Epidemiologists increasingly use information from *validation studies* to correct for the attenuation in estimates of effect induced by measurement error (1–3). Quite often, however, the standard to which the usual measure is compared is itself measured with error. For example, dietary histories might be compared with weighed food records, or self-reported prescription drug use might be compared with medical records. But the weighed food records are clearly not perfect measures of usual diet and the documentation that a prescription was written does not prove that the drug was purchased and used.

The purpose of this paper is to explore the influence of the errors in an *alloyed* gold standard when a validation study is used to correct estimates of effect. We show below that commonly used correction techniques

generally will not correct perfectly. Instead, they can "overcorrect" or "undercorrect" to produce an inflated or a reduced estimate of effect. The direction and magnitude of the bias in the corrected estimate depend on the sizes of the errors in the alloyed gold standard and in the usual measure and on the correlation between the errors. The magnitudes and correlations of these errors cannot be estimated unless a true gold standard is available.

We begin with a hypothetical example of a 2 × 2 table with a misclassified dichotomous exposure and a superior but imperfect measure of exposure used for validation. We then develop the theory in the case of continuous covariates to show how the direction and amount of bias depend on the extent of the measurement error in the putative gold standard and its correlation with the errors in the usual measure. We use this to calculate the impact of possible errors in the standard used for a study to validate diet histories.

TWO BY TWO TABLES

Improper corrections can be obtained when using validation studies for dichoto-

TABLE 1. Results of hypothetical validity study using X, Z, and W in 200 nondiseased and 200 diseased subjects

	Truly exposed (X = 1)				Truly unexposed (X = 0)			
	Cases		Controls		Cases		Controls	
	Z = 1	Z = 0	Z = 1	Z = 0	Z = 1	Z = 0	Z = 1	Z = 0
W = 1	48	12	24	6	8	32	12	48
W = 0	32	8	16	4	12	48	18	72

mous covariates. Consider a situation with a case-control study where W is the measured exposure variable and a validation study with measurements of W, Z, and X, where Z is more sensitive and specific than W and where X is the true value. Assume that W and Z are nondifferential misclassifications of the exposure X. Table 1 presents the most likely outcomes of validation studies in 200 nondiseased subjects and 200 diseased subjects from the base, assuming that Z and W have specificities and sensitivities of 80 percent and 60 percent (with X as the standard), respectively, and that the misclassifications in Z and W are independent. Tables 2, 3, and 4 contain panels with hypothetical 2 × 2 tables from this data set.

The matrix correction method (3, 4) using X as the gold standard and sensitivity and specificity of W both estimated as 0.6 corrects the attenuated odds ratio estimate of 1.2 to 3.0, as it should. But, if Z, the better, but still misclassified measure, is used as the gold standard, the apparent sensitivity and specificity (using nondiseased subjects only) will be 36/70 = 0.51 and 76/130 = 0.58, respectively. Applying the correction gives an estimated odds ratio of 11, substantially overshooting 3.

When the misclassifications in Z and W are positively correlated, applying the standard correction can also result in *under*correction. For example, consider a new classification W' that equals W with probability 0.5 and Z otherwise. The true specificity and sensitivity of W' are 0.7 and the apparent sensitivity and specificity, based on nondiseased subjects only, are 53/70 = 0.76 and 103/130 = 0.79. The odds ratio based

TABLE 2. Hypothetical 2 × 2 table based on table 1 with exposure classified correctly (X)

	Measured exposure		Total
	Present	Absent	
No. of cases	100	100	200
No. of controls	50	150	200
Observed odds ratio	3.0	1.0	

TABLE 3. Hypothetical 2 × 2 table based on table 1 with exposure misclassified nondifferentially (Z*)

	Measured exposure		Total
	Present	Absent	
No. of cases	100	100	200
No. of controls	70	130	200
Observed odds ratio	1.9	1.0	

* 80% specificity and sensitivity.

TABLE 4. Hypothetical 2 × 2 table based on table 1 with exposure misclassified nondifferentially (W*)

	Measured exposure		Total
	Present	Absent	
No. of cases	100	100	200
No. of controls	90	110	200
Observed odds ratio	1.2	1.0	

* 60% specificity and sensitivity.

on W' is 1.5; correction by the matrix method (4) increases this to 2.2, still below the true value of 3. More strongly associated misclassifications will result in more pronounced undercorrection, with the extreme case $W = Z$ implying perfectly correlated errors, appearing to be completely accurate, and resulting in no correction to the odds ratio.

CONTINUOUS COVARIATES

In this section, we use standard theory for the problem of errors in continuous predictor variables in standard regression. Our intent is to identify a problem rather than calculate its

effects precisely. Results are obtained by application of the definitions of regression and the laws of covariance. Using asymptotic theory, we present large sample properties for errors in continuous predictors in standard regression rather than for the situation of misclassified categories. We find the continuous situation simpler and a better aid to intuition. Our results apply to corrections in proportional hazards regression models with continuous covariates, including Cox regression (5), and logistic regression models for cohort (6) and case-control data (7), at least approximately when the distributions of the true covariates and errors are normally distributed.

We denote the perfectly measured variable as X with variance σ_X^2 and the outcome variable as Y, where $E(Y|X) = \beta_0 + \beta_X X$. Instead of measuring X, which might be difficult or expensive to obtain or unknown, an approximation or available proxy W is used in the main study, with

$$W = X + e_W, \tag{1}$$

errors e_W independent of X, $E(e_W) = 0$, $var(e_W) = \sigma_W^2$, and e_W contributing no information about Y beyond what is already in X, that is, nondifferential error or $cov(e_W, Y|X) = 0$. The slope β_W of the regression of Y on W is related to the true slope β_X of the regression on X by

$$\beta_X = \frac{\beta_W}{R_{XW}},$$

where

$$R_{XW} \equiv \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2} = \frac{var(X)}{var(W)}$$

is the reliability of W as a measure of X (2). A simple estimate of R_{XW} is the square of r_{xw}, the correlation of X and W from a validation study. With the estimate r_{xw}^2, and the estimate $\hat{\beta}_W$ based on W we can obtain an estimate, corrected by a validation study of W and X, $\hat{\beta}_{W(X)}$ of β_X (2):

$$\hat{\beta}_{W(X)} = \frac{\hat{\beta}_W}{r_{xw}^2}.$$

Thus, $\hat{\beta}_{W(X)} = \beta_X$, or, more precisely, since we are assuming a large sample, $\hat{\beta}_{W(X)}$ is a consistent estimate of β_X.

We now assume that an alloyed gold standard measure Z is used in place of X in the validation study, with

$$Z = X + e_Z, \tag{2}$$

$E(e_Z) = 0$, $var(e_Z) = \sigma_Z^2$, and $cov(e_Z, Y|X) = 0$. The errors e_Z and e_W may have correlation $\rho_{ZW} \equiv \sigma_{ZW}/(\sigma_Z\sigma_W)$, where the covariance between e_Z and e_W is $cov(e_Z, e_W) = \sigma_{ZW}$.

We define $\hat{\beta}_{W(Z)}$ as the estimate obtained by correcting β_W using the squared correlation r_{zw}^2 from the validation study with the alloyed gold standard instead of using r_{xw}^2, i.e.,

$$\hat{\beta}_{W(Z)} = \frac{\hat{\beta}_W}{r_{zw}^2}. \tag{3}$$

To find the relation between $\hat{\beta}_{W(Z)}$ and β_X, we note that:

$$var(Z) = \sigma_X^2 + \sigma_Z^2,$$

$$var(W) = \sigma_X^2 + \sigma_W^2, \tag{4}$$

$$\hat{\beta}_W = \beta_W = \beta_X \frac{\sigma_X^2}{\sigma_X^2 + \sigma_W^2},$$

and

$$r_{ZW}^2 = \frac{(\sigma_X^2 + \sigma_{ZW})^2}{(\sigma_X^2 + \sigma_Z^2)(\sigma_X^2 + \sigma_W^2)}.$$ (5)

After substituting in expression 3 for β_W and r_{ZW}^2 with equations 4 and 5, respectively, we can obtain

$$\hat{\beta}_{W(Z)} = \beta_X \frac{\sigma_X^2(\sigma_X^2 + \sigma_Z^2)}{(\sigma_X^2 + \sigma_{ZW})^2}$$ (6)

$$= \beta_X \frac{\sigma_X^2(\sigma_X^2 + \sigma_Z^2)}{[\sigma_X^2 + \rho_{ZW}/(\sigma_Z\sigma_W)]^2}$$

$$= \beta_X \frac{1 + \sigma_Z^2/\sigma_X^2}{(1 + \rho_{ZW}\sigma_Z\sigma_W/\sigma_X^2)^2}.$$ (7)

Thus, $\hat{\beta}_{W(Z)}$, the estimate corrected by a validation study with the alloyed Z, is not in general a consistent estimate of β_X.

The relation between the corrected estimate and the true slope depends on whether the correlation in the errors of Z and W is zero, positive, or negative. If errors in the alloyed Z and the available measure W are independent, then $\sigma_{ZW} = \rho_{ZW} = 0$,

$$r_{ZW}^2 = r_{XZ}^2 r_{XW}^2,$$ (8)

and $\hat{\beta}_{W(Z)} > \beta_X$, implying overcorrection. With $\sigma_{ZW} \neq 0$, $\hat{\beta}_{W(Z)}$ may be greater or less than β_X, depending on the relative values of σ_W^2 and σ_{ZW}. A positive correlation between errors e_Z and e_W ($\sigma_{ZW} > 0$), which seems more likely than a negative one, will diminish the overcorrection, leading to perfect correction when $(\sigma_X^2 + \sigma_{ZW})^2$ equals $\sigma_X^2(\sigma_X^2 + \sigma_Z^2)$ (equation 6), or:

$$\rho_{ZW} \equiv \frac{\sigma_{ZW}}{\sigma_Z\sigma_W} = \frac{\sigma_X\sqrt{(\sigma_X^2 + \sigma_Z^2)} - \sigma_X^2}{\sigma_Z\sigma_W},$$ (9)

and undercorrection for greater σ_{ZW}. A negative correlation between errors will accentuate the overcorrection.

TABLE 5. Ratio of corrected regression coefficient to true regression coefficient (extent of over- or undercorrection) due to use of an alloyed gold standard

ρ_{ZW}^*	$\sigma_W^2/\sigma_X^2 = 1$†			$\sigma_W^2/\sigma_X^2 = 0.5$†		
	σ_Z^2/σ_X^2‡			σ_Z^2/σ_X^2‡		
	0.5	0.25	0.1	0.5	0.25	0.1
−0.75	6.80	3.20	1.89	3.84	2.31	1.59
−0.50	3.59	2.22	1.55	2.67	1.84	1.39
−0.25	2.21	1.63	1.30	1.96	1.50	1.23
0	1.5	1.25	1.1	1.5	1.25	1.1
0.25	1.08	0.99	0.94	1.19	1.06	0.99
0.50	0.82	0.80	0.82	0.96	0.90	0.89
0.75	0.64	0.66	0.72	0.79	0.78	0.81
Correlation yielding perfect correction§	0.32	0.24	0.15	0.45	0.33	0.22

* Correlation between errors in alloyed standard (Z) and approximate measure (W).
† Ratio of variances of error in approximate measure (W) and in gold standard (X).
‡ Ratio of variances of error in approximate measure (W) and in alloyed gold standard (Z).
§ Based on equation 9.

Table 5 shows the extent of over- or undercorrection (ratio of corrected regression coefficient to true one) based on equation 7 for different values of σ_Z^2 and σ_W^2 relative to σ_X^2 and of ρ_{ZW} under this scenario. The central row in the table corresponds to the situation where the mechanisms underlying the errors in Z and W are independent. When the errors in the imperfect measures are uncorrelated and the variance of Z exceeds the variance of X by 50 percent, the bias factor will be 1.5. In general, the bias factor depends on the amounts of error in Z and W relative to the variability in X in the study population, as well as the correlation between the errors in Z and W. As ρ_{ZW} increases, the bias factor falls, equaling 1 at the correlation shown in the final line. Thus, when the errors are negatively correlated, independent, or weakly positively correlated, the corrected estimate will overcorrect beyond the true value. Undercorrection occurs when there is a moderate or strong positive relation between the errors.

So far we have shown that $\hat{\beta}_{W(Z)}$ is not a consistent estimator of β_X. In general, it is not even a consistent estimate of β_Z, the slope of the regression on the alloyed standard Z. To see this, note first that

$$\beta_X = \beta_Z \frac{\sigma_X^2 + \sigma_Z^2}{\sigma_X^2}.$$

Upon substitution in equation 7, we obtain

$$\hat{\beta}_{W(Z)} = \beta_Z \frac{(\sigma_X^2 + \sigma_Z^2)^2}{(\sigma_X^2 + \sigma_{ZW})^2}.$$

Thus, $\hat{\beta}_{W(Z)} = \beta_Z$ only if $\sigma_{ZW} = \sigma_Z^2$, for instance, when W is equal to Z plus an *independent* error, viz.,

$$W = Z + e_W^*, \tag{10}$$

with $\text{cov}(e_W, e_W^*) = \text{cov}(X, e_W^*) = 0$, or, equivalently, $E(e_W^*) = 0$ and $\text{cov}(Z, e_W^*) = 0$. This result is not surprising since here Z plays the role of X in the simple measurement error model 1.

Example

Errors in measuring diet for a case-control study of cancer of the colon and rectum (8, 9) were investigated in a validation and reliability study (7, 10). We concentrate here on the comparison between dietary histories, the instrument used in the main study, and 30-day weighed-food records, assumed to generate a more accurate estimate of usual 24-hour diet. The correlation between the estimates of fat consumed daily (measured in units of 100 calories) obtained by the two methods was 0.53. If one assumes that the food records are the gold standard (X in our notation), then the correction factor in equation 1 is $(1/0.53)^2 = 3.56$, which takes the estimated logistic regression coefficient of 0.05, based on the fat consumption estimated from dietary history (W), and corrects it to $0.05 \times 3.56 = 0.18$. These corre-

spond to a crude estimate of an odds ratio of 1.65 associated with consumption of 1,000 extra calories from fat, and a corrected estimate of 5.93.

However, while food records may depict usual diet more accurately than the histories, it would be quite optimistic to assume that the food records perfectly reflect usual diet and hence constitute a gold standard. If the errors in the weighed food records (now denoted by Z while the unknown gold standard is denoted by X) entailed errors uncorrelated to the errors in the diet histories, the correction factor of 3.56 would be too great. For example, independent errors in Z and W ($\sigma_{ZW} = 0$) and values of $\sigma_X^2 = 4.7$, $\sigma_Z^2 = 0.7$, and $\sigma_W^2 = 9.7$ are compatible with the validation study results (7, 10). In this scenario, Z is a much better approximation to X than W is, yet still has 15 percent more

variation than X ($[\sigma_X^2 + \sigma_Z^2]/\sigma_X^2 = 1.15$), implying that the naive correction factor of 3.56 is too big by a factor of 1.15. Thus, a more appropriate correction factor would be $3.56/1.15 = 3.09$, yielding a corrected coefficient of $0.05 \times 3.09 = 0.15$ and an estimate of the relative increase in risk associated with an additional 1,000 g of fat of 4.70.

Correlation in the errors of Z and W can lead to over- or undercorrection, or to perfect correction. Without the assumption of $\rho_{ZW} = 0$, a wide range of values of σ_X^2 and ρ_{ZW} are compatible with the results of the validation study. The values $\sigma_X^2 = 4.1$, $\sigma_Z^2 = 1.3$, $\sigma_W^2 = 10.3$, and $\rho_{ZW} = 0.16$ imply that the original correction is appropriate. A smaller correlation implies a smaller overcorrection, while a larger correlation implies undercorrection. Values of $\rho_{ZW} = 0.36$, $\sigma_X^2 = \sigma_Z^2 = 2.7$, and $\sigma_W^2 = 11.7$ lead to undercorrection by a factor of 0.66.

DISCUSSION

We have shown that using the correlation coefficient from a validation study to correct for attenuation can lead to important bias when the "gold standard" is not perfect. Surely, less than perfect gold standards are the norm for occupational, environmental, and nutritional exposures. That one measure is clearly superior to another does not make the better one a gold standard, and some correction methods can go awry in common situations when a fallible measure is wrongly assumed to be a gold standard. Epidemiologists ought to be more cautious about accepting error corrections, even those based on validation studies. Use of a poor substitute for the measure of interest in the validation study, no matter how much better than the available measure, can inappropriately inflate the estimate of effect obtained from correction. For example, using medical records to validate self-reported prescription drug use, as in a study of exogenous estrogens as a risk factor for endometrial cancer, may not be the gold standard it is often thought to be. A recent unpublished report (A Survey on the Need for a Prescription Drug Benefit under the Medicare Pro-

gram. American Association of Retired Persons, Washington, DC) estimates that 27 percent of Americans above age 44 years either do not fill their prescriptions at all or do not take any of the medication. Thus, an apparently low accuracy of the self-report may result from errors in the records rather than in the self-reports.

Data from a validation study can be used to correct for attenuation by methods other than the one we have described above that uses the correlation coefficient. Methods proposed include correcting by $\text{var}(W)/[\text{var}(W) - \text{var}(W - X)]$ where $\text{var}(W)$ is estimated from the main study and $\text{var}(W - X)$ is estimated from the validation study (7, 11) and by the slope of the regression of X on W from the validation study (6). Estimating $\text{var}(W)/\text{var}(X)$ as $\hat{\sigma}_w^2/\hat{\sigma}_x^2$ directly from the validation study also seems natural. Each method gives a different corrected estimate when X is an alloyed standard, but all are, in general, biased. The approach of Rosner et al. (6) appears to be the most robust against use of alloyed instead of actual gold standards, as it does give an unbiased estimate when the errors in the two measures being compared in the validation study are independent.

We propose no solution to the general problem of estimating β_X or β_Z from validation data with an alloyed gold standard. In particular, while use of expression 7 may suggest a consistent estimator provided relevant parameters can be estimated, we fear that poor small sample properties outweigh any asymptotic advantages it might have.

We considered the standard or classical error model (12) for reasons of familiarity and simplicity, but others are possible. For example, the value of the error could be correlated with the true value X (11). We see no reason to believe that correction methods for other models are immune from the problems we discuss.

The direction and magnitude of the correlation of possible errors between the measures studied and the actual error in the presumed gold standard (equation 7) influence the bias in the corrected estimate. It is important to consider the likely direction of

the correlation in designing and interpreting validation studies, since these quantities cannot be known or estimated without a measured gold standard. Positive correlation between the errors is likely when two measures tend to make the same mistakes. Someone who underreports ice cream consumption on a food frequency questionnaire may also avoid (or underreport) consumption of ice cream while completing a diary, and, thus, both instruments could yield similar error with respect to a gold standard of "usual" diet. The scenarios in table 5 suggest that quite small correlations between errors can turn overcorrections due to alloyed gold measurements into undercorrections. On the other hand, independent errors seem likely when two different sources are used, such as self-report and medical or dental records (where errors might be due to sloppiness or treatment elsewhere) for history of diagnostic x-rays. Negative correlations seem less likely but can also occur. A negative correlation could be induced if the true protective factor for breast cancer were amount of green vegetables (X), while the validation study was used to compare amount of any vegetable (W) to amount of cruciferous vegetables (Z), which was mistakenly taken to be the true risk factor, when consumption of yellow vegetables (the major part of $W - X$) is positively correlated with consumption of noncruciferous green vegetables (the major part of $-(Z - X)$). In general, negative correlations are likely when W is too sensitive and Z is too specific for X, or when W is too specific and Z is too sensitive for X.

We sometimes wish to estimate the coefficient β_Z, on the alloyed measure, rather than β_X, on the true value. For instance, following an improvement in technique for measuring a prognostic indicator, we might wish to update a "prognostic index" regression model to allow prediction using the better, albeit imperfect, tool. Here it is important to distinguish between model 10, under which a simple correction is appropriate, and model 1, under which correction is more complex. Model 10 requires that the available measure W be the sum of the im-

proved measure Z and an independent error. We suspect that plausible error mechanisms rarely correspond with model 10, although it may sometimes be a reasonable approximation. With a sufficiently rich validation study, it may be possible to determine the correct measurement error model and thereby estimate β_Z without bias.

Correction methods can work even without a gold standard in reliability studies when an assumption of independent errors is made. Freedman et al. (13) have described how one can correct for the related problem of variation over time in reported diet: if errors in each of several temporally spread records are independent, one can estimate an effect without bias. Similarly, methods have been developed for correcting for misclassification when there is more than one assessor of exposure (14, 15); however, Brenner (16) has pointed out that the relation between the errors affects the corrected estimates, just as association between misclassification of approximate measures and alloyed gold standards affects the correction in a validation study.

Corrections for attenuation often are used informally in discussion. For example, one aspect of a review by Poole and Trichopoulos (17) of the evidence of associations between exposure to magnetic fields and cancer discussed the impact of errors in measuring exposure. They note that wiring code configuration, which has been used as a proxy measure of residential exposure to magnetic fields, has a low correlation (between 0.4 and 0.6) with another proxy, spot measurements. They argue that, assuming $r = 0.55$, a relative risk of 2 based on the proxy would reflect a true relative risk of $9.9 = \exp[\log(2)/0.55^2]$. The authors acknowledge that one step in this argument is the assumption that the correlation between the two proxy measures Z and W is close to the correlation between a proxy and the etiologically relevant index. Our work suggests caution in extrapolation of this sort. For example, if the errors in the two proxies are uncorrelated, then the correlations between X and Z and between X and W could each be 0.74, which is the square root of

0.55 (see equation 8). This implies a correction of an apparent relative risk of 2 based on either proxy to a true one of only $3.5 = \exp[\log(2)/0.74^2]$, much more likely to be missed in ecologic studies of temporal or geographic contrasts associated with electricity use than the 9.9 that they suggested.

Validation studies serve three important purposes. They permit comparison of a practical measurement instrument against a superior but less practical measure. They can spur improvement of instruments. Finally, when the better instrument is extremely good, they can be used to "correct" relative risk estimates. The potential for serious bias depends in a complex way on the properties of the instruments being compared, particularly the variances of the true measure, the alloyed gold standard, and the available measure as well as the correlation between the errors. It is sobering to realize that substantial bias can occur under plausible conditions. More research is needed to identify when it is realistic to expect that a corrected estimate will be more useful than an estimate based on the usual measure alone.

REFERENCES

1. Willett W. An epidemiologic perspective on exposure measurement error. Stat Med 1989;8: 1031–65.
2. Armstrong BG. The effects of measurement errors on relative risk regressions. Am J Epidemiol 1990; 132:1176–84.
3. Willett W. Nutritional epidemiology. New York: John Wiley & Sons, 1989.
4. Barron BA. The effects of misclassification on the estimation of relative risk. Biometrics 1977;33: 414–18.
5. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 1982;69:331–42.
6. Rosner B, Willett WC, Spiegelman D. Corrections of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989;8:1051–69.
7. Armstrong BG, Whittemore AS, Howe GR. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. Stat Med 1989;8:1151–63.
8. Jain M, Davis GM, Grace FG, et al. A case-control study of diet and colo-rectal cancer. Int J Cancer 1980;32:757–68.
9. Miller AB, Howe GR, Jain M, et al. Food items and food groups as factors in a case-control study of diet and colo-rectal cancer. Int J Cancer 1983; 32:155–61.
10. Jain M, Howe GR, Johnson KC, et al. Evaluation of a diet history questionnaire for epidemiologic studies. Am J Epidemiol 1980;111:212–19.
11. Snedecor GW, Cochran WG. Statistical methods. 7th ed. Ames, IA: Iowa State University Press, 1980;
12. Carroll RJ. Covariance analysis in generalized linear measurement error models. Stat Med 1989;8: 1075–93.
13. Freedman LS, Carroll RJ, Wax Y. Estimating the relation between dietary intake obtained from a food frequency questionnaire and true average intake. Am J Epidemiol 1991;134:310–20.
14. Hui SL, Walter SD. Estimating error rates of diagnostic tests. Biometrics 1980;36:167–71.
15. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol 1988;41: 923–37.
16. Brenner H. Use and limitations of dual measurements in correcting for nondifferential exposure misclassifications. Epidemiology 1992;3:216–22.
17. Poole C, Trichopoulos D. Extremely low-frequency electric and magnetic fields and cancer. Cancer Causes Control 1991;2:261–76.